

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:47:23 ; Search time 210.42 Seconds  
(without alignments)  
6.825 Million cell updates/sec

Title: US-09-331-631a-8\_COPY\_120\_161

Perfect score: 245  
Sequence: 1 SQRFQECQCHQHQEQRP.....QCVRECKRYQENPMRGER 42

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_36.\*  
1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	245	100.0	590	19 W62832	Gossypium hirsutum
2	111.5	45.5	525	19 W62831	Theobroma cacao an
3	111.5	45.5	566	13 R20181	Sequence encoded b
4	99	40.4	625	19 W62830	Macadamia integrif
5	98	40.0	666	19 W62829	Macadamia integrif
6	94	38.4	666	19 W62828	Macadamia integrif
7	74	30.2	28	19 W62841	Stenocarpus sinuat
8	70	28.6	637	19 W62837	Hordeum vulgare an
9	66	26.9	816	16 R71111	Spirocerbellar at
10	66	26.9	816	20 Y33494	Human SCAL protein
11	64	26.1	1898	20 Y30795	A human trichohyal
12	63	25.7	71	20 Y09181	Peptide seq ID No:

13	63	25.7	184	12 R14193
14	63	25.7	184	15 R54981
15	63	25.7	184	16 R76694
16	63	25.7	184	18 W15415
17	63	25.7	184	19 W30526
18	63	25.7	184	21 Y49926
19	63	25.7	189	9 P81140
20	63	25.7	189	15 R54980
21	62.5	25.5	35	13 R21079
22	62.5	25.5	175	14 R33390
23	62	25.3	86	20 W95073
24	62	25.3	86	20 W95078
25	62	25.3	94	20 W95075
26	62	25.3	94	20 W95080
27	62	25.3	108	20 W95071
28	62	25.3	108	20 W95076
29	62	25.3	184	19 W30670
30	62	25.3	910	20 Y22191
31	61.5	25.1	33	19 W62816
32	61.5	25.1	593	19 W62816
33	61.5	25.1	1211	19 W47028
34	61	24.9	425	12 R13792
35	61	24.9	678	14 R42087
36	60.5	24.7	409	20 W90342
37	60.5	24.7	489	20 W90341
38	59	24.5	140	10 P91891
39	59	24.1	371	20 W73369
40	58	23.7	395	17 W03474
41	58	23.7	737	16 R76640
42	58	23.7	737	16 R76639
43	58	23.7	737	18 W18317
44	58	23.7	2074	21 Y54319
45	57	23.3	214	20 Y06303

## ALIGNMENTS

RESULT 1	
W62832	ID W62832 standard; Protein: 590 AA.
XX	AC W62832:
XX	27-OCT-1998 (first entry)
XX	Gossypium hirsutum antimicrobial protein.
XX	antimicrobial protein; Infestation: control.
XX	Gossypium hirsutum.
XX	OS
XX	PN W09827805-A1.
XX	PD 02-JUL-1998.
XX	PE 22-DEC-1997; 97WO-AU00874.
XX	PF 20-DEC-1996; 96AU-0004275.
XX	PR
XX	PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX	Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
XX	WPI, 1998-377279/32.
XX	Novel anti-microbial protein from e.g. Macadamia integrifolia -
XX	useful for controlling microbial infestations of plants or mammals
XX	Claim 1; Page 49-51; 96pp; English.
XX	The sequence is that of an antimicrobial protein which can
CC	be used to control microbial infestations in plants and mammalian





[illegible]

PT	useful for controlling microbial infestations of plants or mammals
XX	
PS	Claim 1; Page 60-62; 96pp; English.
XX	
CC	The sequence is that of an antimicrobial protein which can
CC	be used to control microbial infestations in plants and mammalian
CC	animals.
XX	
SQ	Sequence 637 AA;
<hr/>	
Query Match	28.6%; Score 70; DB 19; Length 637;
Best Local Similarity	30.6%; Pred. No. 1.6;
Matches 15; Conservative	7; Mismatches 13; Indels 14; Gaps 2
OY	6 DECOOHCHDGEORPEKKOQCVCNECKEYDEN-----FMARGE 41
	I I
Ddb	42 qgcvcvrcfcrdeiprr-yshacvcqecrdddqgrmrhcgceeqgrgrrgwge 89

```

RESULT      9
R71111
ID   R71111 standard; Protein; 816 AA.
XX
XX   R71111;
XX
XX   22-SEP-1995 (first entry)
XX
DE   Spinocerebellar ataxia type 1 (SCA 1) protein product.
XX
XX   Spinocerebellar ataxia type 1; SCA 1; presymptomatic diagnosis.
XX
XX   Homo sapiens.
XX
XX   W09501437-A.
XX
XX   12-JAN-1995.
XX
XX   29-JUN-1994; 94WO-US07336.
XX
XX   29-JUN-1993; 93US-0084365.
XX
XX   28-JUN-1994; 94US-0267803.
XX
XX   (MINU ) UNIV MINNESOTA.
XX
XX   Chung M, Orr HT, Zoghbi HY;
XX
XX   WPI; 1995-061001/08.
XX
XX   N-PSDB; Q84793.
XX
XX   New autosomal dominant spinocerebellar ataxia type 1 nucleic acid
XX
XX   used to develop prods. for detection or presymptomatic
XX
XX   diagnosis of a SCA1 disorder
XX
XX   Claim 31; Fig 15; 11pp; English.
XX
XX
XX   Q84793 is a new autosomal dominant spinocerebellar ataxia type 1
XX
XX   (SCA1) nucleic acid, it encodes the protein product described in
XX
XX   R71111. Both the nucleic acid and the protein can be used to develop
XX
XX   products, for the presymptomatic detection of a SCA 1 disorder.
XX
XX
XX   Sequence      816 AA;

Query Match      26.9%; Score 66; DB 16; Length 816;
Best Local Similarity 52.2%; Pred. No. 5.9;
Matches 12; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

OY      2 QROPECCQHCHQOQOREPEKKQ 24
db      200 qqqqqqqqqhqqqqqqqqqqqq 222

```



```

PR      02-SEP-1997;       97US--0922201.
XX
PA      (SEQU-) SEQUENOM INC.
PI
Higgins GS, Koester H, Little D, Lough D:
DR      WPI: 1999-312406/26.
N-PSDB; X37239.
XX
Mass spectrometry detection of polypeptides useful for determining
PT genetic predisposition to Huntington's disease and prostate cancer
XX
Example 1; Page 130; 134pp; English.
XX
The invention relates to a process for determining the identity of a
CC target polypeptide, using mass spectrometry to determine the molecular
CC mass, and comparing this with molecular mass of a known polypeptide. The
CC method comprises (a) obtaining the polypeptide by in vitro translation,
CC or transcriptions then translation, of a nucleic acid encoding the
CC polypeptide; (b) determining the molecular mass of at least one of the
CC obtained fragments by mass spectrometry; and (c) comparing the fragment'
CC molecular mass with that of the fragments of a corresponding known
CC polypeptide. The methods are used to detect the presence or disposition
CC for genetic diseases or conditions particularly those caused by an
CC abnormal number of trinucleotide repeats in a gene, such as Huntington's
CC disease, prostate cancer, Fragile X syndrome type A, myotonic dystrophy
CC type I, Kennedy disease, Machado-Joseph disease, and dentatorubral or
CC pallidolysian atrophy, spinobulbar muscular atrophy or ageing, and also
CC for genotyping, in forensic analysis and parental testing, where the
CC nucleotide repeats, preferably di-, tri-, tetra-, or penta-nucleotide
XX repeats, are quantified.
XX
Sequence       71 AA:
SQ
Query Match          25.7%; Score 63; DB 20; Length 71;
Best Local Similarity 46.2%; Pred. No. 1.2;
Matches    12; Conservative   7; Mismatches     7; Indels    0; Gaps    0
OY      2 ORQFECCQHCHQOEORPEKKQCVR 27
        |:|:|::|||::|:|:|:|
Db      17 qqqgqggqhghqggqggqggqlsr 42
RESULT  13
ID      R14193 standard; Protein; 184 AA.
AC      R14193:
DT      11-DEC-1991 (first entry)
DE      Napin 1-2 gene product.
XX
Desaturase; expression cassette; B. campestris.
FN      WO9113972-A.
PD      19-SEP-1991.
PF      14-MAR-1991; 91MO-U001746.
PR      14-NOV-1990; 90US-0615784.
PR      16-MAR-1990; 90US-0494106.
PR      13-AUG-1990; 90US-0567373.
PA      (CALG-) CALGENE INC.
PI      Thompson G, Knauf V;
DR      WPI: 1991-295627/40.
N-PSDB; Q13969.
XX

```

```

PT DNA encoding a plant desaturase used for modifying the satd.
PT fatty acid compsn. of plant cells and plant seeds
XX
XX Disclosure: Fig 10; 128bp; English.
PS
CC An expression construct contg. a plant desaturase suitable for plant
CC transformation is produced. 5' upstream and 3' downstream sequences
CC obtainable from the B. campestris napin gene here are used.
CC See also Q1363-69.
XX
S0 Sequence 184 AA;

Query Match 25.7%; Score 63; DB 12; Length 184;
Best Local Similarity 25.4%; Pred. NO. 3;
Matches 18; Conservative 7; Mismatches 8; Indels 38; Gaps 3
OY 2 ORQFOE-----COOHCHQO-----EORPEKKQOCV 26
Db 48 rkeftgagahlkacqgmlhkqamsgspswtldgfdfeddvengqgpgqrp11lqgc- 106
:::|::|::|::|::|
OY 27 RECREREYQENP 37
Db 107 --cnehhqgeep 115
|::|::|::|

RESULT 14
R54981
ID R54981 standard; Protein; 184 AA.
XX
XX R54981;
AC
XX
XX 07-DEC-1994 (first entry)
DT
XX
XX Napin protein from Brassica campestris.
DE
XX
XX enhance seed oil production; acyl carrier protein; ACP-I.
KW
XX
XX Brassica campestris.
OS
XX
XX US5315001-A.
PN
XX
XX 24-MAY-1994.
PD
XX
XX 31-JUL-1986; 86US-0891529.
PE
XX
XX 31-JUL-1986; 86US-0891529.
PR
XX
XX 28-JUL-1986; 86US-0891529.
PR
XX
XX 15-NOV-1989; 89US-0437764.
PR
XX
XX 28-JAN-1992; 92US-0826696.
XX
XX (CALJ ) CALGENE INC.
FA
XX
XX Knauf VC, Kridl JC;
PI
XX
XX WPI: 1994-166646/20.
DR
XX
XX N-PSDB: Q65428.
XX
XX DNA encoding plant acyl carrier protein - used to enhance prodn.
PT of seed oil and to modulate the fatty acid compsn. of the oil
PT
XX
XX Disclosure: Fig 4E-4G; 28bp; English.
XX
XX R54981 shows a Brassica napin protein, generated from lambda CGN1-2
CC clone. The napin gene promoter is regulated during seed maturation.
CC This can be used to control the expression of an ACP in seed to
CC enhance the production of seed oil and modulate the fatty acid
CC composition. The DNA can also be used to prepare probes.
XX
XX Sequence 184 AA;
S0

```

Best Local Similarity 25.4%; Pred. NO. 3;  
Matches 18; Conservative 7; Mismatches 8; Indels 38; Gaps 3;

Matches 18; Conservative 7; Mismatches 8; Indels 38; Gaps 3;

Search completed: March 1, 2001, 15:47:24  
Job time: 249 sec

Job time: 249 sec

QY	2	ORGRQEE-----COOHCHQ-----SOREEKKQGC	26
Db	48	rkefegqahlkacqgwhlkqamsgsgspwldgefafeddvenngqsgppqgrppllgc-	106
QY	27	RECREKYOENP	37
Db	107	--chelhgeep	115

RESULT 15  
P7669A

R76694 standard; Protein; 184 AA.

AC R76694;

DT 04-NOV-1995 (first entry)

DE B. campestris napin.

KW Napin, promoter; seed-specific gene transcription; oilseed;  
KW transgenic plant.

OS *Brassica campestris*.

PN US5420034-A.

PD 30-MAY-1995.

PF 31-JUL-1986; 86US-0891529.

PR	09-JUL-1990;	90US-0550804.
PR	31-JUL-1986;	86US-0801539

PR	28-JUL-1987;	87US-0078538.
PR	25-JAN-1988.	88US-0147781

PR 08-AUG-1991; 91US-0742834.  
XX

PA (CALJ ) CALGENE INC.  
XX

PI knauf VC, Kridl JC;  
XX

DR WPI: 1995-206250/27.  
DR N-PSDB: 094586.

XX	DNA providing seed
PT	

PT transcription initiation  
seed cells, e.g. the  
PT

Pt and Brassica plants.

ps Disclosure; Fig. 2A-J; 40pp; English  
XX

CC	Clone Lambda CGN1-2
CC	Library of B. campest

CC CDS as well as exten  
CC sequenced (094586)

CC the 3' - and 5' -ends  
XX

Query Match	25.7%; Score 63; DB 16; Length 184;
Best local similarity	95.4% (best vs 3)

